IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant	: Stephen K. Pinto et al.	Art Unit	: 2123
Serial No.	: 10/826,630	Examiner	: Juan Carlos Ochoa
Filed	: April 16, 2004	Conf. No.	: 1302
Title	: PREDICTIVE MODEL GENERATION		

Mail Stop Appeal Brief - Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

PRE-APPEAL BRIEF REQUEST FOR REVIEW

Claim 1[1]

The rejection of claim 1 based on Cabena[2], Harrison[3], and AAPA[4] is clearly wrong, because not one of the references, alone or in combination, described or would have made obvious any of the following five claimed actions used in generating a predictive model:

1. Selecting variables having at least a first predetermined level of significance from a pool of predictor variables to form a first population of predictor variables.

2. Extending the first population of predictor variables to include cross products of the variables in the first population.

3. Selecting variables in the extended first population that have at least a second predetermined level of significance to form a second population of predictor variables.

4. Selecting at least one variable having less than the first predetermined level of significance from the pool of predictor variables (recited in action 1) to produce cross products. The cross products are included in the second population to form an extended second population of variables.

5. Selecting variables in the extended second population and having at least a third predetermined level of significance to form a third population of predictor variables.

Cabena described none of the five actions. Contrary to the examiner's assertion that Cabena described the first action in the second and third paragraphs of page 101, those parts of

---

[1] The applicant has proposed formal amendments. The applicant's argument does not rely on the amendments.
[2] Intelligent Miner for Data Applications Guide.
[3] An Intelligent Business Forecasting System.
[4] Crisp-DM 1.0 step-by-step data mining guide.

Cabena described visualization of the modeling results, not selecting variables for generating a predictive model. In addition, although Cabena did describe determining "whether there are strong variables, ... split the data into multiple files ...", Cabena did not "select variables having at least a first predetermined level of significance", but instead ran "RBF against *each* of the separate files..." (third paragraph, page 101). In other words, although Cabena may have identified whether there are strong variables, he used all variables for display and no selection was done.

Also contrary to the examiner's assertion that Cabena partially described the second action by expanding his population of variables using supplementary variables, Cabena's supplementary variables are only used for profiling, not for defining the clusters (first paragraph, page 48). Cabena referred to: "... supplementary variables to aid in the *interpretation* of the neural cluster" (last paragraph, page 52), "... supplementary variables, which are profiled by model region but *not* used to *build* the model" (second paragraph, page 100), and "input all data as supplementary not used in the prediction" (last paragraph, page 118). Cabena did not expand the population of predictor variables for use in generating a predictive model.

The examiner conceded that Cabena did not describe the other three actions of claim 1 listed above. But the examiner alleged that AAPA described "cross products" and—in combination with Cabena—described the second action referred to above. And the examiner contended that AAPA described actions 3 through 5.

AAPA did not describe and would not have made obvious actions 3 through 5. With regard to action 3, although AAPA constructed new attributes from cross products of existing attributes and extended the population of the attributes (page 50, "Derived Attributes"), AAPA did not select "variables having at least a second predetermined level of significance from the extended first population of predictor variables" after including the cross products.

With regard to action 4, AAPA did not describe subsequent actions of including more cross products of variables into "the second population of predictor variables" after previous actions of including cross products. AAPA's general description of data construction using cross products would not have made obvious the multi-action variable selection and predictor variable population extension of claim 1. In addition, AAPA did not describe and would not have made obvious that at least one of the variables producing the cross product has "less than the first

predetermined level of significance", also recited in action 3. AAPA listed data to be used or excluded based on selection criteria (page 48, "Output"), but said nothing about whether those data excluded would be used for constructing cross products (page 50, "Output").

AAPA also did not select "variables having at least a third predetermined level of significance from the extended second population of predictor variables" after including the cross products. AAPA said nothing about selecting variables after including cross products.

Harrison had nothing to do with the actions of selecting variables and extending a population of predictor variables for generating a predictive model, recited by claim 1.

Accordingly, the combination of Cabena, Harrison, and AAPA would not have made obvious the features of claim 1.

Claim 31

The rejection of claim 31 based on Cabena, Harrison, and Galperin[5] is also clearly wrong, at least because none of the references, alone or in combination, described displaying to a user "concordance scores".

The examiner interpreted "concordance scores" to be "area under curve". However, claim 31 explains that concordance scores are obtained based on "a receiver-operator-characteristic curve".

The examiner conceded that Cabena and Harrison did not describe displaying to a user "concordance scores" but alleged that Galperin did. Galperin calculated integral criterion of lift to maximize the lift within a range (abstract, column 3, lines 18-20, and column 4, line 7), which had nothing to do with "concordance scores". As claim 31 explains, concordance scores indicate "to the user goodness of fit of the historical data to the generated predictive model"; while lift as a term of art in data mining, is a measure of the performance of a model in segmenting the population. The lift of a subset of the population is defined as the predicted response rate for that subset divided by the predicted response rate for the population. For example, suppose a population has a predicted response rate of 5%, but a certain model has identified a segment with a predicted response rate of 20%. Then that segment would have a lift of 4.0 (20%/5%) (http://en.wikipedia.org/wiki/Lift_(data_mining)). Lift or integral criterion of lift has nothing to

[5] US6,640,215.

do with concordance scores that indicate to the "user goodness of fit of the historical data to the generated predictive model".

Accordingly, Galperin did not calculate concordance scores and did not display to a user "concordance scores". The combination of Cabena, Harrison, and Galperin did not describe and would not have made obvious the features of claim 31.

Claim 34

The rejection of claim 34 based on Cabena, Harrison, and AAPA is also clearly wrong, at least because none of the references, alone or in combination, described or would have made obvious combining two models "based on response propensities of each model" in order to create cross-modal deciles, let alone based on "weaving of the historical data" to provide cross-modal optimization or concatenating the predictions of the two models.

The examiner conceded that Cabena had nothing to do with combining models. The examiner also conceded that although Harrison described combining models, Harrison did not describe and would not have made obvious combining two models "based on response propensities of each model" in order to create cross-modal deciles, let alone based on "weaving of the historical data" to provide cross-modal optimization or concatenating the predictions of the two models.

The examiner alleged that AAPA described combining based on response propensities in order to create cross-modal deciles and based on data weaving to provide cross-modal optimization. However, the examiner failed to point out which part of AAPA described such features. In fact, AAPA had nothing to do with combining models. Nothing was combined based on "response propensities," and no data was woven. AAPA said nothing about "to create cross-modal deciles" or "to provide cross-modal optimization". AAPA mentioned "propensity" (page 41 "Data mining success criteria"), but only as a selected criteria for determining success of a project.

Accordingly, none of Cabena, Harrison, and AAPA, alone or in combination, described or would have made obvious the features of claim 34.

The dependent claims are patentable over the cited references, for at least the same reasons discussed with respect to the independent claims, from which they depend.

Please apply any charges or credits to deposit account 06-1050, referencing attorney docket 17146-0007001.

Respectfully submitted,

Date: 3/7/9

David L. Feigenbaum
Reg. No. 30,378

Fish & Richardson P.C.
225 Franklin Street
Boston, MA 02110
Telephone: (617) 542-5070
Facsimile: (877) 769-7945